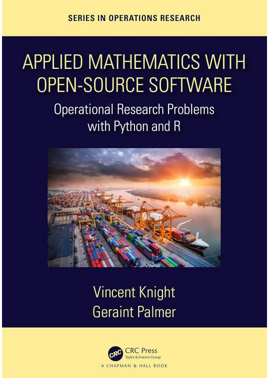
Cooperation

Vince Knight

2024-03-04 until 2024-03-29


SERIES IN OPERATIONS RESEARCH
APPLIED MATHEMATICS WITH OPEN-SOURCE SOFTWARE
Operational Research Problems with Python and R
Vincent Knight
Geraint Palmer
CRC Press
Taylor & Francis Group
A CHAPMAN & HALL BOOK


NASH STREET

Golden Balls

Watch the final stage of a game show.

Golden Balls

Watch the final stage of a game show.

Who expected that outcome?

I really enjoy this video as it plays to our gender stereotypes.

Who played correctly? Who was right?

If honor was important then that implies the value of the money is not everything here. If we just consider the money then she played correctly.

An implicit reason for honor as a value is that it creates reputation. This is something we are going to explore through game theory.

Definition

$$A = \begin{pmatrix} R & S \\ T & P \end{pmatrix} \qquad B = \begin{pmatrix} R & T \\ S & P \end{pmatrix}$$

with the following constraints:

$$T > R > P > S \qquad 2R > T + S$$

Definition

$$A = \begin{pmatrix} R & S \\ T & P \end{pmatrix} \qquad B = \begin{pmatrix} R & T \\ S & P \end{pmatrix}$$

with the following constraints:

$$T > R > P > S \qquad 2R > T + S$$

- The first constraint ensures that the second action "Defect" dominates the first action "Cooperate".
- The second constraint ensures that a social dilemma arises: the sum of the utilities to both players is best when they both cooperate.

This game is a good model of agent (human, etc) interaction: a player can choose to take a slight loss of utility for the benefit of the other play **and** themselves.

The Nash equilibria is for both individuals to defect.

So if this is the case: how come we see so much cooperation around?

Definition

Given a two player game $(A, B) \in \mathbb{R}^{m\times n^2}$, referred to as a stage game, a T-stage repeated game is a game in which players play that stage game for $T > 0$ repetitions. Players make decisions based on the full history of play over all the repetitions.

Definition

Given a two player game $(A, B) \in \mathbb{R}^{m\times n^2}$, referred to as a stage game, a T-stage repeated game is a game in which players play that stage game for $T > 0$ repetitions. Players make decisions based on the full history of play over all the repetitions.

As an example you could start to wonder what would happen if the participants of that game show were sharing 20% of their winnings five times in a row.

It would not be necessarily clear that individuals should steal in the first turn.

Definition

A strategy for a player in a repeated game is a mapping from all possible histories of play to a probability distribution over the action set of the stage game.

Definition

A strategy for a player in a repeated game is a mapping from all possible histories of play to a probability distribution over the action set of the stage game.

In the case of the split or steal game repeated 5 times. How many strategies are there?

Number of possible actions

Number of repetitions

$$2^{\sum_{i=0}^{4} 2^i \times 2^i}$$

History of play of first player

History of play of second player

$$= 4{,}479, 489, 484, 355, 608, 421, 114, 884, 561, 136, 888, 556, 243, 290, 994, 469, 299,$$
$$069, 799, 978, 201, 927, 583, 742, 360, 321, 890, 761, 754, 986, 543, 214, 231, 552$$

$$> 10^{102}$$

So how do we study such a thing?

$$\sum_{i=0}^{\infty} \delta^{i} u_i(s_1, s_2)$$

Probability of game continuing

$$\sum_{i=0}^{\infty} \underset{\text{Probability of game continuing}}{\delta^{\,i}} u_i(s_1, s_2)$$

One approach is to use an infinitely repeated game with a discounting factor δ. This is a common approach in mathematics when dealing with tricky problems.
In our case if we restrict ourselves to 3 strategies:

- Always cooperate s_C;
- Always defect s_D;
- Grudger: cooperate until defected against and then defect forever s_G

Reward for mutual cooperation

$$U(s_C, s_C) = U(s_G, s_G) = U(s_G, s_C) = U(s_C, s_G) = \sum_{i=0}^{\infty} \delta^i R = \frac{R}{1-\delta}$$

$$U(s_D, s_D) = \sum_{i=0}^{\infty} \delta^i P = \frac{P}{1-\delta}$$

Punishment for mutual defection

$$U(s_D, s_G) = T + \sum_{i=1}^{\infty} \delta^i P = T + \frac{P\delta}{1-\delta}$$

Temptation to defect

$$U(s_C, s_C) = U(s_G, s_G) = U(s_G, s_C) = U(s_C, s_G) = \sum_{i=0}^{\infty} \delta^i \overset{\text{Reward for mutual cooperation}}{R} = \frac{R}{1-\delta}$$

$$U(s_D, s_D) = \sum_{i=0}^{\infty} \delta^i \underset{\text{Punishment for mutual defection}}{P} = \frac{P}{1-\delta}$$

$$U(s_D, s_G) = \underset{\text{Temptation to defect}}{T} + \sum_{i=1}^{\infty} \delta^i P = T + \frac{P\delta}{1-\delta}$$

We can now see if deviation from mutual cooperation is rational. This is the case if and only if:

$$\begin{aligned}
\frac{R}{1-\delta} &< T + \frac{P\delta}{1-\delta} \\
R - T &< \delta(R + P - T) \\
\delta &< \frac{R - T}{R + P + T} && \text{assuming large } T
\end{aligned}$$

Thus for large enough value of T as long as the probability of the game ending (δ) is small enough then deviation is rational. Conversely, if δ is large enough than cooperation is rational.

This indicates that one secret to cooperation is long interactions.

Robert Axelrod[1] (1943 -)
Courtesy of University of Michigan personal website,
https://commons.wikimedia.org/w/index.php?curid=20096037

[1] Robert Axelrod. "Effective Choice in the Prisoner's Dilemma". In: *The Journal of Conflict Resolution* 24.1 (1980), pp. 3–25. (Visited on 03/20/2024).

Robert Axelrod[1] (1943 -)
Courtesy of University of Michigan personal website,
https://commons.wikimedia.org/w/index.php?curid=20096037

[1] Robert Axelrod. "Effective Choice in the Prisoner's Dilemma". In: *The Journal of Conflict Resolution* 24.1 (1980), pp. 3–25. (Visited on 03/20/2024).

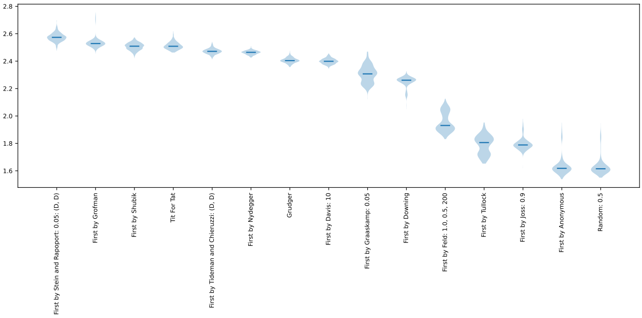
Robert Axelrod created a computer tournament inviting people[2] to submit strategies by computer code to play 200 turns of the Prisoners Dilemma with

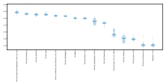
$$A = \begin{pmatrix} 3 & 0 \\ 5 & 1 \end{pmatrix}$$

14 strategies and a completely random on were submitted. The simple strategy Tit For Tat one: this strategy starts by cooperating and then just mimics the opponents last action.

Axelrod ran a second tournament[3] immediately afterwards, 64 strategies were submitted. Everyone knew the result of the first tournament and nevertheless Tit For Tat won again.

Axelrod wrote a widely cited book[4] describing this and the implications of the work.

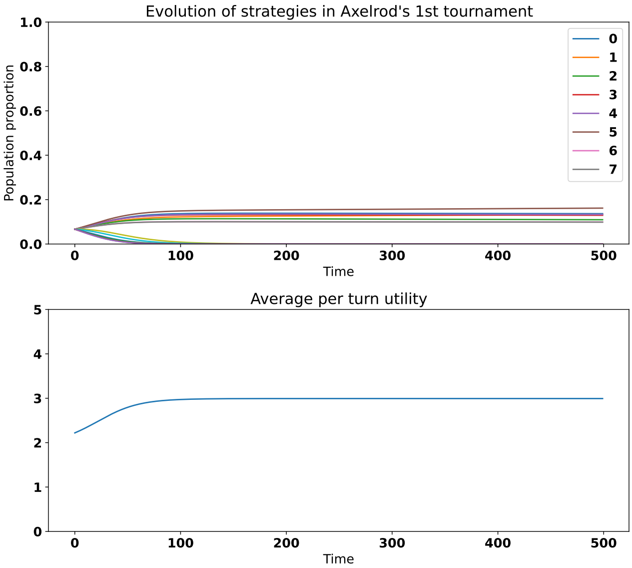

2.8
2.6
2.4
2.2
2.0
1.8
1.6
First by Stein and Rapoport: 0.05: (D, D)
First by Grofman
First by Shubik
Tit For Tat
First by Tideman and Chieruzzi: (D, D)
First by Nydegger
Grudger
First by Davis: 10
First by Graaskamp: 0.05
First by Downing
First by Feld: 1.0, 0.5, 200
First by Tullock
First by Joss: 0.9
First by Anonymous
Random: 0.5



Using an open source library which was starting in Namibia in 2015 but has since become a popular and incredibly research tool (with over 230 strategies) it is possible to reproduce the tournament.

Actually that's not true, it's possible to fail to reproduce the tournament. The general conclusions are somewhat the same though.

If we put the results of this tournament in the replicator dynamics equation we can see the overall behaviour that emerges


Evolution of strategies in Axelrod's 1st tournament
Population proportion
Time
0
1
2
3
4
5
6
7
Average per turn utility
Time

Looking at the plot on the left, we see that there are only a few strategies that survive the evolutionary process. In fact those that do are ones that only cooperate against each other.

- **Be nice**
- **Be provocable**
- **Don't be envious**
- **Don't be too clever**

- Be nice
- Be provocable
- Don't be envious
- Don't be too clever

These four properties still hold even though the research is not necessarily reproducible.
This was used for a long time to justify the conditions for the emergence of cooperation.

- **Be nice**: do not defect first.
- **Be provocable**: reciprocate both cooperation and defection.
- **Don't be envious**: focus on your happiness as opposed to making sure you are happier than your co player.
- **Don't be too clever**: by scheming to exploit the opponent

"The world of game theory is currently on fire."

MIT Technology Review

"The world of game theory is currently on fire."

MIT Technology Review

This was a commentary on a paper from Press and Dyson[5].
They proved a theorem that showed the there exists a strategy that can always extort their co player.

For $p = \left(\overset{P(C\,|\,CC)}{p_1}, \overset{P(C\,|\,CD)}{p_2}, \underset{P(C\,|\,DC)}{p_3}, \underset{P(C\,|\,DD)}{p_4}\right)$ and $q = (q_1, q_2, q_3, q_4)$:

$$u(p,q) = \begin{vmatrix} -1 + p_1q_1 & -1 + p_1 & -1 + q_1 & f_1 \\ p_2q_3 & -1 + p_2 & q_3 & f_2 \\ p_3q_2 & p_3 & -1 + q_2 & f_3 \\ p_4q_4 & p_4 & q_4 & f_4 \end{vmatrix}$$

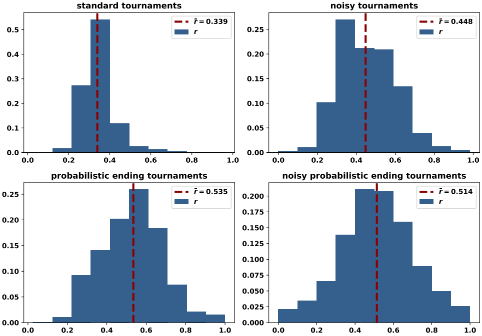
For $p = (\overset{P(C|CC)}{p_1}, \overset{P(C|CD)}{p_2}, \underset{P(C|DC)}{p_3}, \underset{P(C|DD)}{p_4})$ and $q = (q_1, q_2, q_3, q_4)$:

$$u(p, q) = \begin{vmatrix} -1 + p_1 q_1 & -1 + p_1 & -1 + q_1 & f_1 \\ p_2 q_3 & -1 + p_2 & q_3 & f_2 \\ p_3 q_2 & p_3 & -1 + q_2 & f_3 \\ p_4 q_4 & p_4 & q_4 & f_4 \end{vmatrix}$$

They prove a clever results that relates the scores of the two strategies to this determinant using Cramer's rule.
Importantly there are two columns of that determinant that are exclusively controlled by each player.
This ensures (and there is some algebra missed here) that a given player can extort another player.

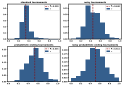
Properties of Winning Iterated Prisoner's Dilemma Strategies[a]

[a] Nikoleta E. Glynatsi, Vincent Knight, and Marc Harper. *Properties of Winning Iterated Prisoner's Dilemma Strategies*. 2024. arXiv: 2001.05911 [cs.GT].



Properties of Winning Iterated Prisoner's Dilemma Strategies[a]

[a]Nikoleta E. Glynatsi, Vincent Knight, and Marc Harper. *Properties of Winning Iterated Prisoner's Dilemma Strategies.* 2024. arXiv: 2001.05911 [cs.GT].



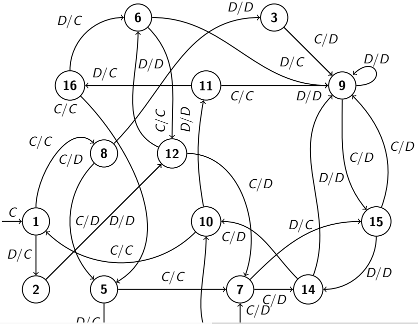
However this doesn't seem to be the case. And indeed the world of game theory is not that on fire.

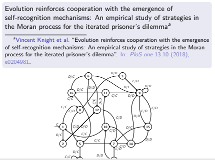
1. Be nice in non-noisy environments or when game lengths are longer
2. Be provocable in tournaments with short matches, and generous in tournaments with noise
3. Be a little bit envious
4. Be clever
5. Adapt to the environment (including the population of strategies).

The last point is particularly important here.

Evolution reinforces cooperation with the emergence of self-recognition mechanisms: An empirical study of strategies in the Moran process for the iterated prisoner's dilemma[a]

[a] Vincent Knight et al. "Evolution reinforces cooperation with the emergence of self-recognition mechanisms: An empirical study of strategies in the Moran process for the iterated prisoner's dilemma". In: *PloS one* 13.10 (2018), e0204981.





An example of this is this strategy here. This is a strategy that is trained using reinforcement learning. It actually recognises itself.

This mechanism of self recognition is one that is commonly seed in evolutionary biology.

nashpy.readthedocs.io
axelrod.readthedocs.io
knightva@cardiff.ac.uk